Some Thoughts On The Virtual Observatory

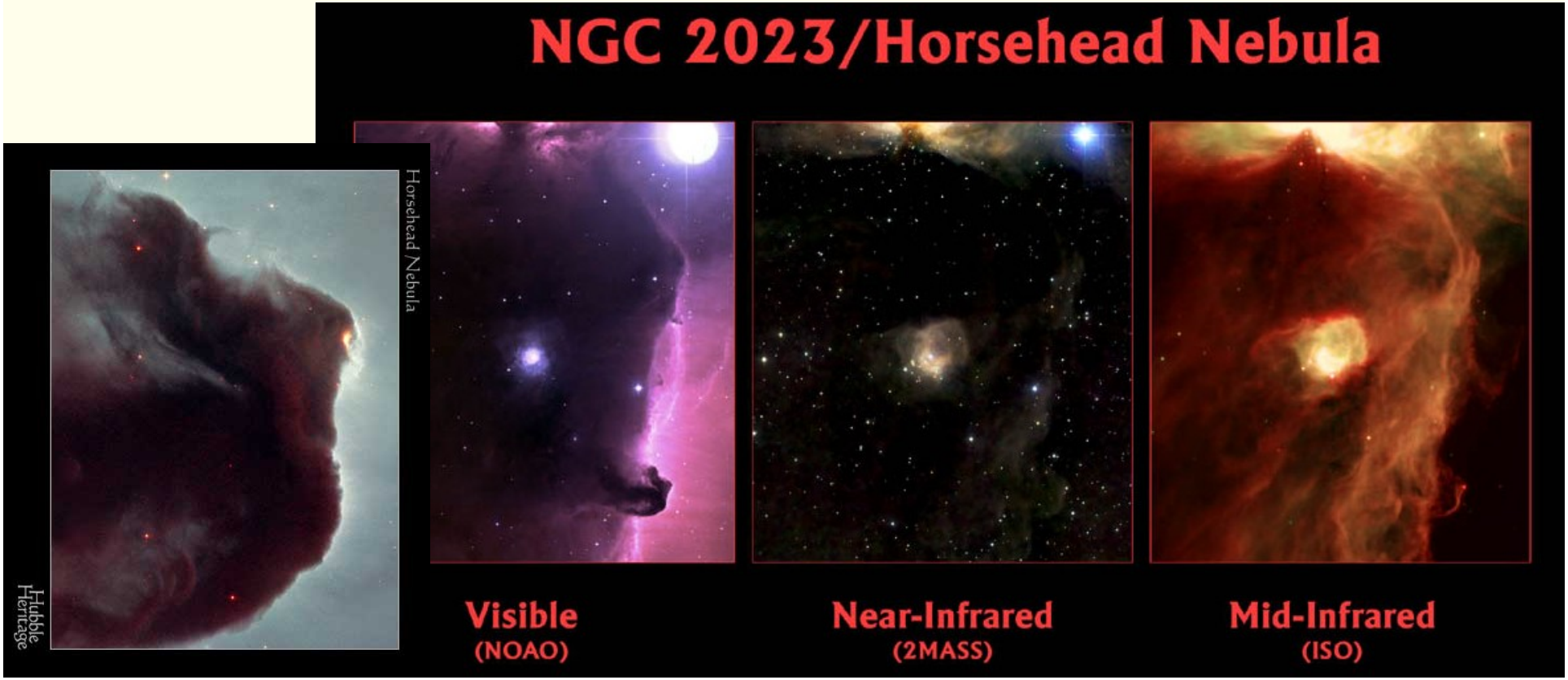
George Helou

ICiS, Park City
August 2011

The Virtual Observatory Vision

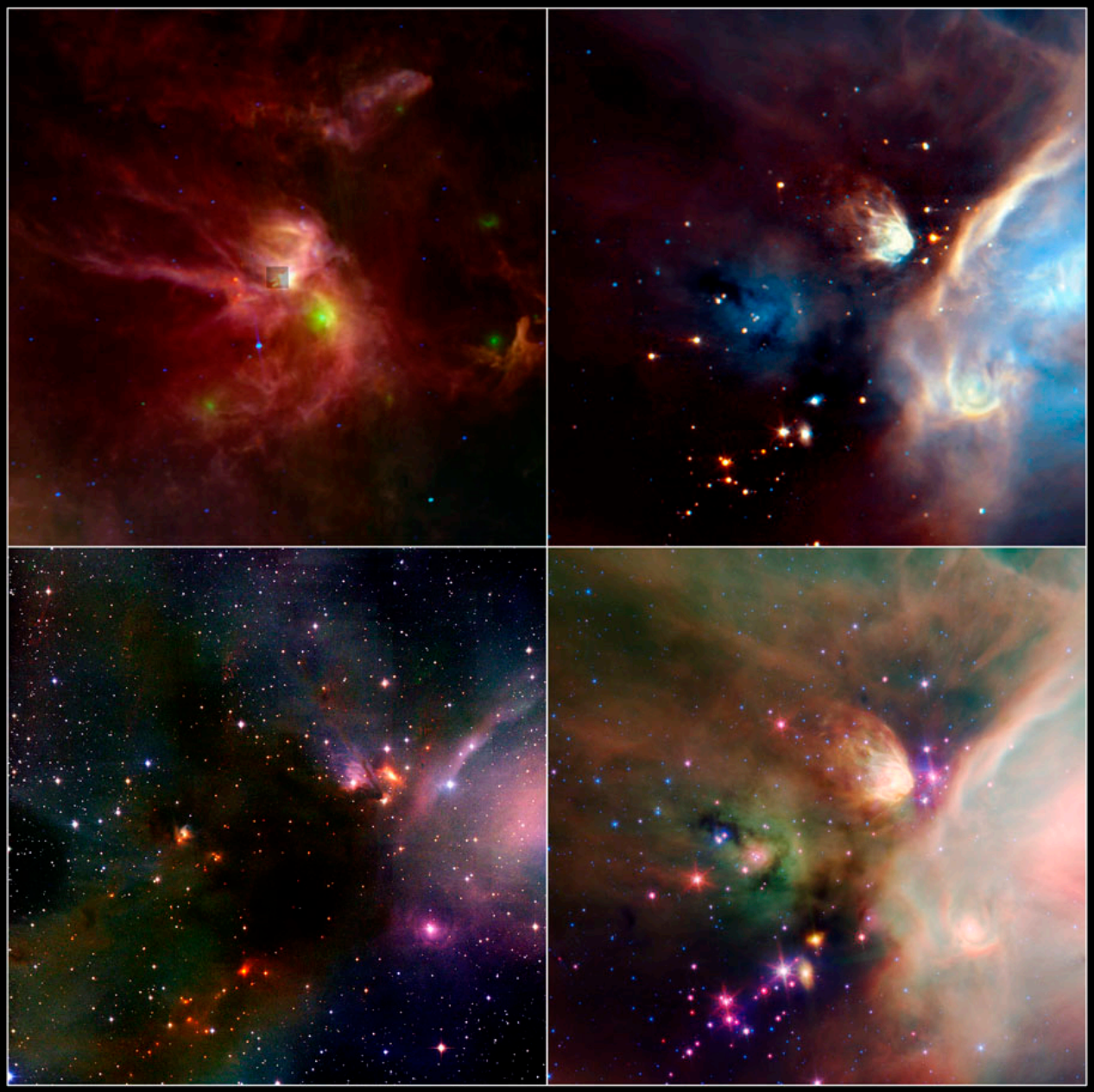
- VO aims to extract new science by combining huge and disparate data sets
  - *Nearly any effectively combined data sets (with $\Delta\lambda/\lambda \geq X$) will offer new science*
  - *Diversity: Among observatories, but also among surveys (eg Kepler vs DES)*
  - *Most of the (observational) data resides at Science Centers and Archives*

- Disclosure: IPAC has been an active participant in NASA & NSF VO projects
  - *Infrared/Submm data sets offer special challenges because of wide range in spatial resolution, data acquisition method, detector properties*



Ambush by Acronym

Acronym	What Do The Letters Mean?	Which is What Exactly?
A VO	Virtual Observatory	General term for a national project providing transparent access to distributed data sets and services through international standards
The VO	Virtual Observatory	All the VO projects operating in celestial harmony as one system
NVO	National Virtual Observatory	U. S. research project to figure out how to make data and services transparently accessible
VAO	Virtual Astronomical Observatory	U. S. operational project to make distributed science services and data access a reality
IVOA	International Virtual Observatory Alliance	International standards body to facilitate international coordination and collaboration among national VO projects



From *Portals to the Universe: The NASA Astronomy Science Centers* (NRC Report, 2007)

"Successful research using archival data sets is dependent on the resident expertise and corporate memory that resides at the science centers."

+ in survey project teams, eg DES or SDSS

Archives, Science Centers and the Virtual Observatory 1

- These are all part of a system – to be optimized globally
- Ideally, simplest path to optimization is hierarchical architecture
  - *Each Wavelength (IRSA, MAST, HEASARC, NRAO, NOAO, etc) and Thematic (NED, ADS) Archive already offers local or domain VO functionality.*
  - *Archives encode their expertise into their local data inventory and retrieval services*
  - *The VAO would be a "thin layer" interacting with each Archive, and through the Archive with its diverse holdings*
- This is one vision, not perfectly aligned the current VAO implementation

Archives, Science Centers and the Virtual Observatory 2

- In this vision…
- VAO would focus on data discovery and protocols, in support of new science services complementing existing offerings
- Archives would focus on creation and serving of data, curation of data, and on science user support

- Roles and responsibilities follow from strengths
  - *Archives (HEASARC, IRSA, MAST, NRAO, NOAO, etc) are all about scientific expertise in mission data sets*
  - *ADS is about powerful access to the literature*
  - *NED is about integration of literature and surveys*
  - *VAO is about connection software, and high-level perspective*

- Responsibility of implementation and maintenance should be distributed among system elements to minimize disruption

Implications

- VO would access data holdings at higher levels set up by the Archives, rather than at the basic DBMS tables (e.g. TAP)
  - *Such access points exist already, and can be easily extended*
- VO standards of near-term value have higher priority, should be implemented in a start-small, extensible form
  - *Archives and VO should partner with community in setting priorities*
  - *Some of the most effective VO standards came out of collaborations among Archives seeking 2-way inter-operability*
- Survey data creators still need to generate their best product and documentation
  - *Implementing full-up, lowest-access-level VO standards can be expensive*

Bottom Line

- "VO" functionality, inter-operability among data sets is here to stay, and is critically important to productivity
- However, it cannot be the panacea for all large data set access/ combination needs
  - *Community, both providers and users, should identify common and near-term needs and communicate those priorities to VO projects*

The Archives are content providers and mission/science experts;
the VAO offers data discovery and protocols to enable new science